The British Academy

Journal of the British Academy, **11(s5)**, 35–58
https://doi.org/10.5871/jba/011s5.035
Posted 14 December 2023

Methodological considerations and assumptions in social science survey research

Daniel B. Wright

Abstract: The articles in this special issue are based largely on results from online social surveys on beliefs and self-reported behaviours related to COVID-19, with an emphasis on ethnicity differences. There are many considerations and assumptions used when conducting this type of research, and when analysing the resulting data, which are often not discussed in the resulting journal articles. These include how the research questions are chosen, how the measurement of the key constructs is done and the analytic approach. The article goes through several of the steps necessary to conduct social science survey research that are often not reported in papers. The aim is to provide a backstage view of how this approach to social scientific questions occurs, pulling back the curtain on these issues.

Keywords: methodology, assumptions, surveys, COVID-19, statistics

Note on the author: Daniel B. Wright is Professor in Educational Psychology at the University of Nevada, Las Vegas. His interests are in quantitative methods and applied cognitive science (in particular, learning in groups).

When COVID-19 began spreading across the world, social scientists began studying the psychological effects of the pandemic on both individuals and societies. They attempted to measure behavioural changes influenced, in part, by the guidelines imposed by health researchers and politicians, and they tried to account for these behaviours using hypothesised psychological constructs and to design interventions both to improve people's psychological well-being and their compliance with health guidelines.

There are several methods that social science researchers use, but the primary method for much social science COVID-19 research has been some form of survey. This is the approach that our group (Rusi Jaspal, Glynis Breakwell, Julie Barnett and myself) has used for most of our research, and some of this is discussed in this special issue. The data discussed are from a British Academy-funded project where we examined the role of identity and other psychological constructs on COVID-19-related beliefs, such as vaccine positivity and behaviours, such as hand washing. Research findings – including our own – are usually disseminated through brief research articles that include a handful of statistics and a couple of figures with arrows connecting the key constructs. This succinct approach to sharing the results, which is efficient for many purposes, can make the decisions involved in conducting such research appear uncontroversial and de-emphasise the considerations and assumptions underlying the approach. The goal here is to focus on the considerations and assumptions. Some findings are reported, but these are only presented as illustrative of the types of research questions addressed. It is rare for the considerations and assumptions to be made explicit in typical journal articles. As such, this article will focus on more methodological concepts and less on psychological theories and their impacts on policies. See the article in this issue, 'Identity resilience, uncertainty, personal risk, fear, mistrust and in-group power influences upon COVID-19 coping' by Glynis M. Breakwell, for a discussion of the theories.

The purpose of this article is to describe, to the broad audience of the *Journal of the British Academy*, the steps that those social scientists who rely on surveys and questionnaires take while conducting research. By making the steps explicit I hope to provide readers with an understanding of not just the research undertaken by myself and my colleagues, but also that of others using this approach. Further, I hope that the article will encourage social science survey researchers to question and to justify their processes/assumptions. The following are, broadly, the key points that our group considered when discussing how to conduct our research and to analyse our data. I use these to structure this article.

1. Create a research team and define research questions/problems to be addressed.
2. Delineate the underlying theories and perspectives that will inform the research.
3. Establish the research design (including population of interest, sampling, *etc*.).

4. Lay out the theoretical constructs to be estimated.
5. Consider the relationships among these constructs.

All of these are interconnected. These are presented as illustrations of the considerations and assumptions that need to be taken into account when undertaking this type of research.

Research team and research questions

Deciding upon the research team and the research questions are often done in concert and affect how the rest of the research pans out.[1] While some research can be undertaken by individuals working alone, it is often useful to have several people with different areas of expertise working together on a project. Sometimes there are people in one's own department who complement your skills, but it is also possible to meet people at conferences or on social media. Our group is composed of senior academics who, through years of experience, have several contacts with complementary skills. A research team may exist and decide what research questions to address, or an individual (or funding organisation) may describe some broad research questions and a team will coalescence around these and fine-tune the questions. In most cases the questions are a combination of researchers' interests and external pressures. Our research team was composed of people with particular social psychological knowledge and different methods skills. We had worked on several projects related to this project together. When the British Academy offered funding on social science research related to ethnicity and COVID-19, our team came together.

Our primary research question for this project, discussed throughout this special issue, was how well identity process theory (IPT) can account for individual differences in vaccine positivity and self-reported likelihood of being vaccinated among different ethnic groups in the United Kingdom and the United States (see the article 'Psychological influences on COVID-19 preventive behaviours and vaccination engagement in the United Kingdom and the United States: The significance of ethnicity' in this special issue). This topic was chosen because of the expertise of our team with respect to this theory and to the methods needed to conduct the social science research. We believe identity resilience is a critical factor that influences people's beliefs and behaviours during a health crisis. This, as discussed in the next section and in detail

[1] I refer to research questions broadly, and simply mean seeking information from the research that will change what the researchers believe on a topic. If a study turns out exactly as expected, the change would be greater certainty in the original beliefs. These might be applied problems that exist in the world, or specific questions that the researchers have concerning a particular theory.

in Breakwell's article 'Identity resilience, uncertainty, personal risk, fear, mistrust and in-group power influences upon COVID-19 coping' in this special issue, is a perspective we bring to the research. The role of identity resilience and other constructs in predicting outcomes is complex and this framework allows us to build a model of these dependencies. It also allows us to examine several research questions simultaneously. Our main research questions are about estimating the size of different relationships.

It is important to note that research groups often come together because participants work in the same department, were at graduate school together, meet at a conference, like the same football team, and so on. Sometimes at large research organisations there is deliberate matching of people, skills, and needs, but even in these settings people often choose who to work on based on how well they get along. Some disciplines and some research questions do not require research teams (e.g., in philosophy and law, articles often have a single author), while others do (e.g., some physics articles will have dozens of authors). If a person's interests are in areas where research teams are useful and they do not have appropriate colleagues in their locale, networking at conferences and on the Internet can help them to find like-minded potential collaborators.

Our perspectives/biases

All research is influenced by the perspectives of the researchers. Francis Bacon (2019 [1620]) recognised this, describing how the *idols* of mind could distort how we interpreted the world. His advice was to avoid these prejudices, so that nature would more truthfully reveal itself. Researchers' beliefs affect how they undertake and interpret research, but completely removing all biases is neither possible nor desirable, as people would then be unable to undertake research or interpret results (e.g., Popper 1994, ch. 4). The legacy of Bacon's desire to be without bias, coupled with the observation that society has achieved some (perhaps much) of his vision of the science-produced industrial society prophesied in *New Atlantis* (Bacon 2020 [1626]), has led to the myth that natural scientists have successfully removed their beliefs, prejudices and biases from the scientific process. In the social sciences, where this myth is less widely believed than in the natural sciences, it seems to have morphed into a desire to emulate natural science, believing that the myth is true for natural scientists, despite the fact that they also succumb to human biases.

One aspect of distancing the researcher from the research is the third-person writing style: 'The author did …' or '*Author's Name* did …', rather than 'I did'.[2] The intent

[2] For papers submitted to blind review this means authors often refer to themselves in the third person when their papers are originally submitted in order to not reveal who they are to the reviewers. Sometimes this third person style is not changed in later versions.

is to distance the research from the researchers. It is not about making the research more objective, but instead props up the illusion that Bacon's goal of removing the idols of mind has been achieved. Norman Campbell (a British physicist and philosopher) questions whether this third-person style is morally correct: 'Here a moral issue is raised. If we are not prepared to make a personal statement in a personal form, are we justified in making it at all?' (Campbell 1928: 1021). Like Bacon, I recognise that people have their beliefs, prejudices and biases and that these affect their research, but like Campbell I believe the formality of much scientific writing, in attempting to discount these influences, is wrong. I will use 'I' when referring to my beliefs/actions and 'we' when referring our group's beliefs/actions. People choose what to study and how to study it, and interpret the observed data using what they already know. Researchers should be both knowledgeable and interested in what they research, and these biases should be influential. But researchers should remain sceptical of any individual findings, particularly their own. Feynman advises scientist that they should be 'bending over backwards to show how you're maybe wrong' (Feynman 1974: 12).

With respect to our group's research, within a survey context we believe that respondents' answers to our survey questions provide information about their attitudes. This implies that they have some level of access to this information. It is known that for many tasks humans do not have conscious access to why they make decisions (e.g., Nisbett & Wilson 1977) and that alternative approaches may be necessary to tap into certain information (e.g., Greenwald *et al.* 1998). Our assumption is that asking people their views, while not having them give precise accurate information about their beliefs, provides responses that are similar enough to their beliefs to be of use. Specifically, we think that responses to questions that we believe a priori are related to a construct can be used to estimate values for each person for the intended construct. Further, we assume that while the individual items will relate to a great many things, taken together the intended construct will be the most prominent of these. These assumptions underlie how much research that uses scales works, and there are psychometric techniques to help evaluate these assumptions.

This does not mean that we believe responses are perfectly accurate. For example, social scientists, referring to COVID-19, often ask behavioural frequency questions, such as: how often do you wash your hands or wear a face mask? Answering questions like this is difficult if respondents are meant to recall each time they did these activities. Surveys usually have respondents choose from a list of either numeric (e.g., one to two times a day) or verbal alternatives (e.g., 'sometimes'); we show that the choice of response alternatives affects estimates and group comparisons (Wright *et al.* 2022). While the numbers that we use to estimate behaviours like hand washing will not be perfectly accurate, we assume that if questions are written appropriately for the sample those people who report more hand washing (or whatever behaviour) will tend to experience this behaviour more than those who report less hand washing; however, we also recognise that systematic biases occur.

Many of the surveys conducted during COVID-19, including our own, have used online survey instruments. There are advantages and disadvantages to this in comparison to other administrative modes. While the sampling is restricted to those who sign up with a company (e.g., MTurk, Prolific), the samples are more representative than, for example, the convenience samples often possible with research on university campuses (Sheehan & Pittman, 2016). Response rate is a problem for all types of surveys. There are problems with all sampling methods and often response rates are low. Pew Research estimates only 6 per cent of the people sampled in its telephone surveys responded in 2018,[3] which also makes estimating non-response bias as difficult. According to the National Research Council (2013) growing non-response rates 'threaten to undermine the validity of inferences obtained through the collection of information from subjects through surveys' (p. x). With online instruments that require respondents to sign up both to the organisation that manages the instrument (e.g., Prolific, Qualtrics, MTurk) and to the specific survey if they see a particular call for that survey in time, a response rate cannot be meaningfully calculated. In much online social science research the interest is in comparing groups that have all been sampled in the same way, either at the same time or over time, and/or exploring associations among people in the sample. Making valid inferences requires assumptions that non-response affects the different groups/times in a similar way.

When online surveys began to become popular there were concerns that respondents would pay less attention to the questions than those taking part in studies in person. However, results show online sample groups often pay more attention than in-person samples, and Prolific samples perform well in comparison with other online sampling methods (Peer *et al.* 2022). Further, the ease of recording response times and click behaviours allows researchers more ways to check the attention paid by respondents than traditional pencil-and-paper surveys.

In addition, it is important to state that our beliefs influence what we choose to examine. Consider our choice to use a 'trust in science' construct. We work in academia and while we hope to be critical about all research, including our own, we believe that the scientific process is better than existing alternatives for allowing wise decision-making, though its error-correction mechanisms could be more efficient. Our belief in the value of the scientific method has influenced our choice to include this construct and our *a priori* belief that it plays an important role in adherence to health guidelines. We trust, to some extent, the scientific literature and rely on this for our

[3] https://www.pewresearch.org/short-reads/2019/02/27/response-rates-in-telephone-surveys-have-resumed-their-decline/ (accessed 15 August 2023).

methods. For example, we use a Trust in Science Scale by Nadelson *et al.* (2014), where they report the psychometric properties for their scale using the data from their sample. The psychometrics they report are specific to their sample and to the time when they administered the scale. As scientists, we remain cautious accepting that any scale measures what the developers say it measures (e.g., Wolff *et al.* forthcoming). A scale does not, for example, have a particular level of reliability. It only has this in relation to a sample. As such, we assume that the scale should have good qualities for our samples, but we do examine this. Later in this article I reveal some of the ways that we did this. That scale is now a decade old and it was created prior to COVID-19. Because we wanted respondents to complete the whole study in a relatively short time, we used a short form of the original questionnaire, composed of six items. Another belief that we have is that IPT provides a good framework to examine the relationships among variables, including identity-related variables. The research described in this special issue is not designed to test IPT as a theory, but to use the theory as a framework to examine various components across respondents from two countries.

Research design

Population and sampling

Details of why a specific population is of interest and why particular sampling approaches are used is not always discussed in empirical papers, so here our rationale is described. Our project aimed to examine differences, by ethnicity, in people in the United Kingdom and United States, applying IPT to COVID-19 beliefs and behaviours. While we are also interested in the relationship among these constructs in other countries, for this project the United Kingdom and United States were our focus. We are interested in the general adult population, meaning all adults above 18, although, as discussed, the sampling procedures mean not all groups were likely to be represented (e.g., with internet surveys, those who seldom use computers will be underrepresented). We did not include people under 18 for two reasons. First, people under 18 in the countries of interest do not have complete authority on whether to have, for example, a vaccine. In most US states parental consent is required for COVID-19 vaccines.[4] In the United Kingdom the situation is slightly more complicated. If a parent does not want their child vaccinated, but the child is judged to be *Gillick competent* (a medical term related to the child being competent to provide consent

[4] https://www.kff.org/other/state-indicator/state-parental-consent-laws-for-covid-19-vaccination/ (accessed 15 August 2023).

without parental consent), the healthcare professional will try to attempt a reconciliation between the child and parent, but the parent cannot overrule a Gillick competent child's decision. The second reason is more pragmatic. Not including people under 18 makes conducting the research simpler as many ethics guidelines require people under that age to provide parental consent to take part in studies.

As noted, the sampling in our studies was done using Prolific and this means the sample is not likely be representative of the population of interest. Not only will those who take part have to have access to the Internet, they will also have to have signed up to Prolific. This means, among other restrictions, that respondents would need to be familiar with Prolific (this probably is why Prolific samples tend to have many current and recent university students as it was created for university research) and would want to take part in research for money. While this is an issue, alternatives during the COVID-19 pandemic were not practical. We thus chose to use an online survey, and we constructed the survey using the popular tool Qualtrics.[5] This allowed a link for the survey to be posted using Prolific,[6] which gave us the option to set up filters (e.g., we wanted UK and US respondents, with quotas that allowed us to have enough respondents in several ethnic categories to allow comparisons). Our survey received ethical approval from the University of Brighton's Cross-School Research Ethics Committee C (Ref: 2022-9564-Jaspal) and all respondents provided consent.

With all surveys, some respondents' data may not be appropriate to use and therefore exclusion criteria exist. Online studies allow some data to be collected that are not available with traditional survey administrative modes. The IP (internet protocol) address is usually available. In our studies we excluded duplicate IP addresses because this may relate to one person who is using two Prolific accounts (as it may not, both accounts are paid for completing the survey). Response times can also be recorded. These can provide a valuable window into the respondent's cognitive processing while answering questions (Luce 1986). Extremely fast responses can indicate that insufficient cognitive processing was done to adequately to answer the questions (see Wise & Kong 2005; Wise 2017, for related discussion). Attention-checking questions, which often include a phrase like 'ignore the rest of this question, just tick option B' (see Gummer *et al.* 2021, for detailed discussion), are often included in both online and other survey formats to check if respondents are reading the questions, but they can confuse some respondents. There are numerous guidelines for constructing online surveys (e.g., Biffignandi & Bethlehem 2021). With online surveys it is possible to force people to respond to each item. If a question is poorly worded or there is some other reason why the respondent feels it is inappropriate for them to respond, this

[5] https://www.qualtrics.com/ (accessed 15 August 2023).
[6] https://www.prolific.co/ (accessed 15 August 2023).

can annoy the respondent and affect the quality of all their subsequent responses or cause them to leave the survey. However, for the types of scales that we use in this study, having people provide an answer for all questions is useful. There are methods to address missing data (e.g., Rubin, 1987; van Buuren, 2018), but if it is believed that each person can provide a meaningful response to an item it is worth having complete surveys.

Which demographics?

The demographics that we were most interested in for the research described in this special issue were ethnicity, gender, age, education, and we also asked some questions about political affiliation. The reasons for this are addressed in the other articles in this special issue. We also asked questions that are of particular importance to the COVID-19 guidelines, including the number of people in a household, because this is related to number of contacts and therefore the possibility of contagion. How demographic questions are asked and which categories are included in the response alternatives can be very contentious. The meaning of, for example, ethnic categories, differs between the United Kingdom and the United States. We tended to follow the ways in which government surveys (e.g., Census Bureau) ask these questions as well as the phrasing suggested in the materials of both our survey programme (Qualtrics) and the sampling program (Prolific). With ethnicity, it is obvious that there is no clear way to differentiate all people and that there is much ethnic variety within any of the categories we choose. For ethnicity, we had Prolific perform a quota sample for the categories it uses for ethnicity. Quota sampling means that you attempt to get a predetermined number of people (a quota) for each category (Kalton 2021). The breakdown we achieved is shown in Table 1. No method for classifying the complexity of ethnicity (or race) adequately captures all the differences.

Table 1. Ethnicity categories for the United Kingdom and United States, as used in the article 'Psychological influences on COVID-19 preventive behaviours and vaccination engagement in the United Kingdom and the United States: the significance of ethnicity' in this special issue.

	UK	**UK%**	**US**	**US%**
Asian	390	35%	111	15%
Black (Black African, Black Caribbean, African American)	388	35%	207	27%
Hispanic, Latino or Spanish of any other origin			180	24%
White (Non-Hispanic)	316	28%	247	32%
Two or More Races/Mixed	17	2%	14	2%
Other	6	1%	4	1%
Total	1117	100%	763	100%

The gender breakdown was 940 (50 per cent) female, 928 (49 per cent) male, 11 (1 per cent) other and there was one missing value. The survey asked respondents for their age in years. Two said they were 9 years old (to have a Prolific account they must be 18 or older, so they likely did not type the first digit), three said they were over 100 (listing their likely year of birth) and one left the age variable blank. Excluding these, the median was 32 years old and the mean was 34.43 years old. The skew towards younger responses is predicted as Prolific began in universities and was seen as a convenient way for students and recent alumni to earn extra money. More details of the demographics are covered in the articles dealing with those. Here the only demographic comparisons were by country: 1117 were from the United Kingdom and 763 were from the United States.

Why R?

There are many statistical packages and no single package is best for all situations. Here we use the free statistical environment R (R Core Team, 2022; for a brief description see Chambers, 2009; for a thorough description see Chambers, 2008). This is one of the most used systems for data analysis; it has been described by Mizumoto & Plonsky (2016) as a *lingua franca* (a shared or bridging language) for both learning and implementing statistics. We used R for this research for at least three reasons. First, it is free, which means anyone can replicate our findings without having to buy expensive software. Second, with over 20,000 free add-on packages and the ability to write your own functions, it allowed us to conduct all the statistical analyses for this project. And finally, this article was written as a document composed of R code for statistical work and LaTeX for word processing, and then these were combined using knitr (Xie, 2015) into a pdf document. One of the concerns about the statistics is not being able to replicate the findings in research reports; this approach allows the finding to be easily replicated (Mair, 2016). The final submitted document can be found on GitHub.[7]

Estimating psychological constructs

Scientists construct models that:

1. They believe approximate nature closely enough to be useful.
2. They believe provide a useful framework to interpret their findings.

Their choice is often influenced by the statistical methods they use, but these statistical methods also influence their theories (Gigerenzer 1991). A popular model that

[7] https://github.com/dbrookswr/BAwork/blob/main/jba2dbwed1.pdf (accessed 15 November 2023).

social science researchers assume is the latent variable model and, as will be clear in this section, this choice relates to both theories and methods. I concentrate on two of the scales discussed in the other articles: six items from the Trust in Science Scale (Nadelson *et al.* 2014) and the sixteen-item Identity Resilience Index (IRI) (Breakwell *et al.* 2022). More details of these are provided in the article 'Psychological influences on COVID-19 preventive behaviours and vaccination engagement in the United Kingdom and the United States: The significance of ethnicity' in this special issue. However, when social scientists use scales in their own contexts the norm is to check at least some of the psychometric qualities of the scale. Because of journal word-length restrictions authors often only give a brief summary of their explorations of the scale.

A latent variable conceptualisation

Latent variable models are taught in both under- and postgraduate social science methods classes. Loehlin and Beaujean (2017) provide an excellent introduction to latent variable models, mathematical details can be found in Bartholomew *et al.* (2011) and Muliak (2010), while Spearman (1904) is a seminal historic text.

An assumption of much social and psychological research is that responses to several related items can be combined to estimate a single construct. For the latent variable model this is because the latent construct is assumed to influence how people answer each of those items. Suppose that you have six variables and believed each is related to a particular construct, say *trust in science*. Figure 1 shows a latent variable model that might be used for this; for the six items we had respondents answer on a 1 to 5 scale from strongly disagree to strongly agree. The arrows mean that what is described in the node at the nock of the arrow *influences* what is described in the node at the arrow's head. The assumption is that responses to each question, for example the *Scientists ignore …* rectangle, are influenced by variation in a respondent's trust in science construct. In addition, responses are also affected by a combination of idiosyncratic aspects of this item and random variation, shown by the *e* nodes to the right of each rectangle. These are often called the error terms associated with the individual items, but it is important to note that they are a combination of error and systematic variation specific to the item. For the model shown in Figure 1, these error terms are assumed to be independent of each other. This means that after taking into account trust in science the variables themselves are independent. There are ways to examine if this assumption is justified, discussed later in this article. In these plots, the most popular convention is to have the latent variables shown in ellipses and the observed variables shown in rectangles.

An important question is whether the latent variable is a dimension, for example from not trusting science at all to trusting science uncritically, or whether the latent

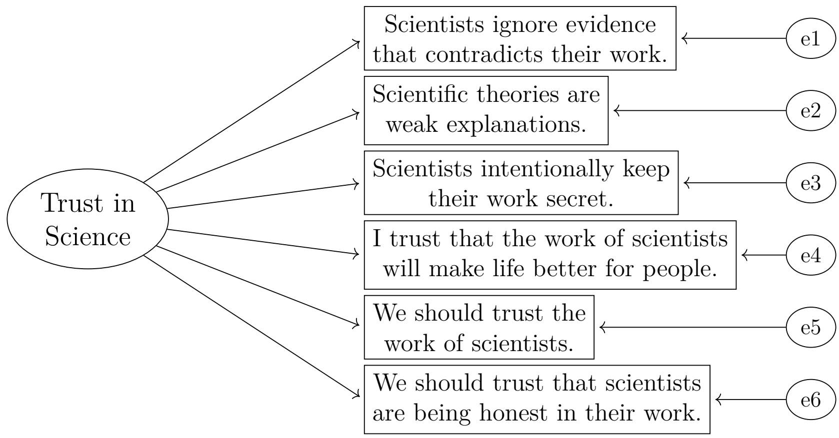

Figure 1. Assumed relationship between a psychological construct, trust in science and six survey items.

variable is categorical, placing people into a small number of groups that share characteristics regarding trust. Latent variable framework allows for both of these characterisations (Bartholomew *et al.* 2011). An approach called taxometric methods (Waller & Meehl 1998) exists that allows researchers to see if the data are more consistent with one or other of these two characterisations, but in many cases the data do not show either being better than the other (Bartholomew 1993). In these cases the researcher chooses what seems most appropriate for their purposes.

When creating scales that other people will use, often those other people will have small samples or even just a few individuals, and will want to create summary measures. This means that using complex methods that require large samples to estimate values for people's constructs may not be practical. Scale designers take this into account and try to design scales where taking the mean of responses, often reverse scoring some of the variables,[8] provides a good estimate for the construct. There are advantages to having a simple method for allowing others to estimate these psychological constructs, and often scales are developed such that the mean of the responses provides a good estimate. This is similar to how teachers report the percentage of correct responses on an assessment for their students. This will probably not have as good statistical properties as more complex procedures (McNeish & Wolf 2020), but in some contexts it is a good option (Widaman & Revelle 2022). The ease of calculating these measures plus the transparency for the students (if Josh and Tommy each get thirty-six questions right, they get the same score) may outweigh other statistical considerations. An often reported measure that is consistent with using the mean of the items to estimate the

[8] The rule for reverse scoring items is if an item goes from m to n, letting $newvar_i = (n + m) - oldvar_i$ means the minimum and maximum possible are the same as the other items.

construct is Cronbach's α, also called Guttman's λ_3 (Guttman 1945; Cronbach 1951). Sometimes this assumption is not valid, but it is still reported. In fact, it is so common reviewers will often ask for it to be reported so that this measure can be compared with others. Several authors have discussed problems with the overuse of and conceptual issues with this measure (e.g., Thompson 2003; McNeish 2018).

Exploring dimensionality with the scree plot

As stressed in the last subsection, the psychometric values of a scale can change over time and with different samples. There are different ways to explore if a scale is measuring the number of constructs that it is intended to measure. I will consider two of the scales that we used in more detail. This illustrates what is done for all scales, but which seldom makes it into articles due to page constraints.

The first measure is the six items from the Trust in Science Scale. Three of the items were reverse scored so that high values on each correspond to more trust in science. The assumption in Figure 1 is that a single latent variable influences all of the observed variables. The assumption is that the scale is unidimensional, with idiosyncratic influences that affect the individual items. This assumption can be examined empirically and it will never be true (i.e., some items will always be more closely related to some other items), but the question is whether the assumption is close enough to be true to be useful. Exploratory data analysis should be conducted, including looking to see if all items are correlated as they should be, prior to creating any latent variables (Wright & Wells 2020). This can be done with both statistical tests, like Pearson's correlations, and visually with scatter plots (see Figure 2). The scatter plots allow outliers to be identified and researchers to check when a straight line seems to describe the relationship well. With typical survey items that are measured on discrete rating scales it is useful to add a small random variable to each point so that each point can be seen. This is called *jittering*. In addition, only 600 of the data points are shown in order to make identifying which coordinates have the most values easier. With most social science applications, the data points are spread out so trying to tell if a pattern is approximately linear is difficult. At this point of the analysis the researcher is usually looking only for clear signs of non-linearity (e.g., is there a floor or ceiling effect) or if the relationship is not monotonic.

Two things can be concluded from these scatter plots. First, there are more responses above 3 (the mid-point on the five-point scale) than below it: 66 per cent compared with 12 per cent. Thus, our sample shows more trust than distrust in science, although there is a spread in responses. Second, the correlations are all at or above $r = .5$. Cohen (1992) describes $r = .5$ as a *large* correlation, so in his terminology all of the associations are large, but his terminology is context dependent. Using Figure 1 as a way of

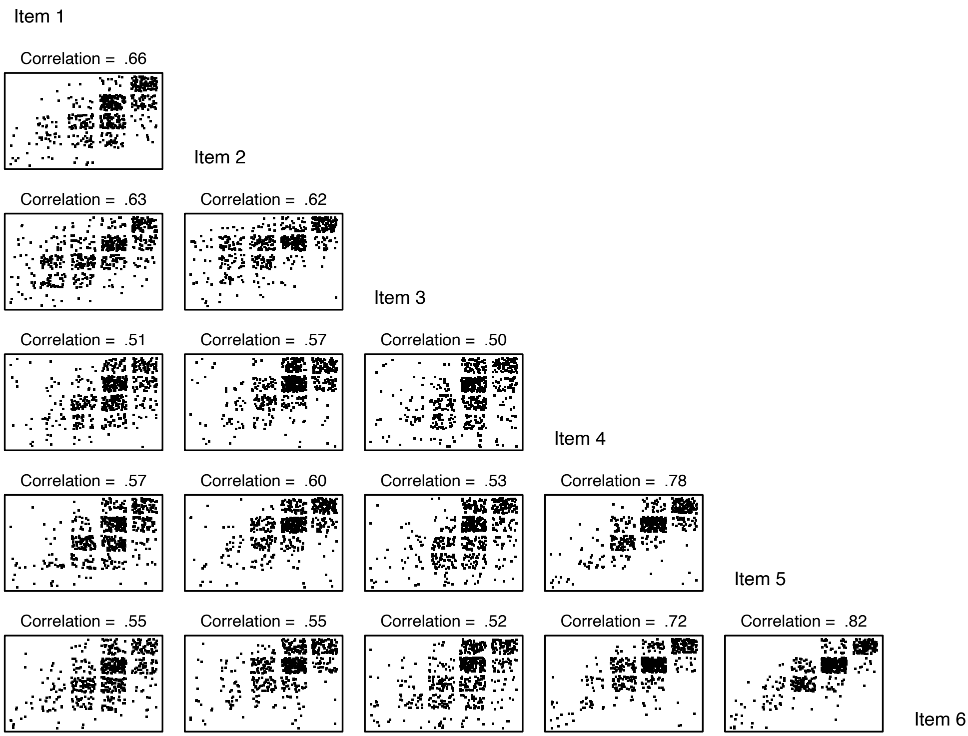

Figure 2. Scatterplot matrix of the six trust in science items. Six hundred cases were randomly chosen and jittered so that it is easier to see the relationships.

describing the size of correlations, suppose that there is a normally distributed latent variable with a standard deviation (SD) of 1 and two observed variables that are this variable plus an item-specific error variable with a SD of 1. Provided that these two error variables are unrelated, the correlation between these two variables would be about $r = .5$. Looking at the spread of the data in the scatter plots also helps one get a feel for what the different values of r mean with respect to how spread out the data are. It is also important to identify any pairs with particularly high or low values.

One of the most used methods to examine the number of dimensions of a set of items is a *scree* test (Cattell 1966). This deserves further explanation as it is often used in a mechanistic way where the researcher just chooses a single value produced by the computer as if this is the 'right' number of dimensions. Scree is the geological term for the loose rubble that has accumulated at the base of a steep hillside. The statistics to construct a scree plot are calculated using the eigenvalues of the correlation matrix. The sum of the eigenvalues of most correlation matrices is the number of variables. The first eigenvalue, which will be the largest, shows how much of this total can be accounted for by a linear combination of the variables or, in lay terms, how much of the variation can be accounted for by a single dimension. The second is how much more can be accounted for by a second dimension and so on. A scree plot is made by drawing a line connecting the eigenvalues. If there are six items, there will be six eigenvalues providing no item is a linear combination of the other five.

The amount accounted for necessarily decreases with each dimension. Cattell likened the underlying structure of a scale to a hillside where the 'scree represents a "rubbish" of small error factors' (Cattell 1966: 249). He describes methods to identify where the scree bends – the elbow – and to use this as the number of dimensions, though he notes that using this approach 'requires the acquisition of *some* art in administering it' (p. 256, emphasis in original).

There are several procedures that can be helpful to guide this art. The most useful in my opinion is adding a line to show how the scree would look if there were no structure to the data. This is called *parallel analysis*. To be part of the hillside you would want to use the dimensions shown on the scree that are well above the random line. What 'well above' means is up to the discretion of the analyst. Velicer *et al.* (2000; see also Auerswald & Moshagen 2019) described several statistical procedures that aim to identify the number of dimensions, and some of these are used later in this section. It is worth noting that reality is much more complex than our models, and that a near infinite number of likely related constructs will inform how people answer any of these questions. Cattell was aware of this: 'There is no such thing as "the true number of factors to extract", since the only possible assumption is that both the number of substantive and the number of error common factors each exceed *n*, the number of variables' (Cattell 1966: 273). The analyst must decide what is appropriate simplification for their purposes to allow them to make what they believe are wise decisions. The scree plots for the trust in science and identity resilience variables are shown in Figure 3, along with lines created to show what the scree would be like for random data. A single eigenvalue stands out above the random scree line for the Trust in Science Scale, but there are several eigenvalues above this line for the Identity Resilience Scale (Breakwell *et al.* 2022).

The second scree plot is for the Identity Resilience Scale, which is discussed in greater detail in other articles in this special issue. It was designed to have four components, and there are four eigenvalues above the line. In most cases like this there are a priori beliefs about the number of dimensions, their meanings and which items each construct will primarily influence. Confirmatory factor analysis (CFA) can be used in this situation. In addition, it is believed that there is still an underlying identity resilience construct that influences all the items, but each item is also influenced by one of the four components (self-esteem, efficacy, distinctiveness and continuity). Thus, the first item can be thought of as:

$$\textit{item 1} = \textit{Identity Resilience} + \textit{Self-Esteem} + e_1 \tag{1}$$

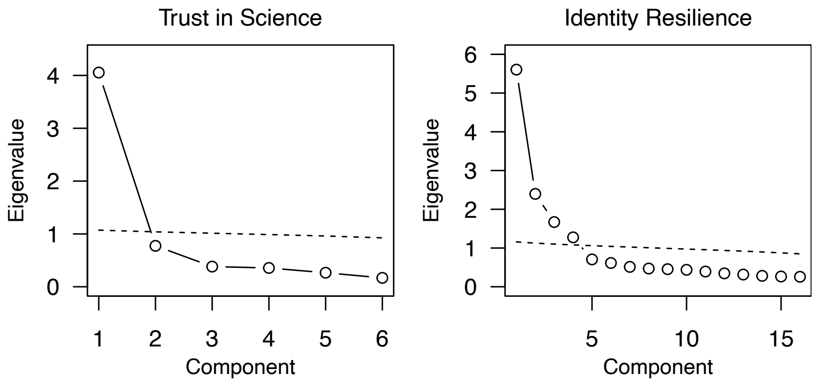

Figure 3. Scree plots for Trust in Science and Identity Resilience Scales.

This is called the bifactor model and is depicted in Figure 4. It shows each of the four proposed components of identity resilience influencing four items (listed on the left side of the figure), and all sixteen of the items being influenced by some overall identity resilience construct. The fit of this model was compared with several alternatives, and this model fits better than the alternatives tested. The choice between exploratory and confirmatory approaches is often difficult. In one sense, exploratory approaches are more data driven, while confirmatory approaches are more guided by, depending on one's perspective, Bacon's idols of the mind or the research questions driving the

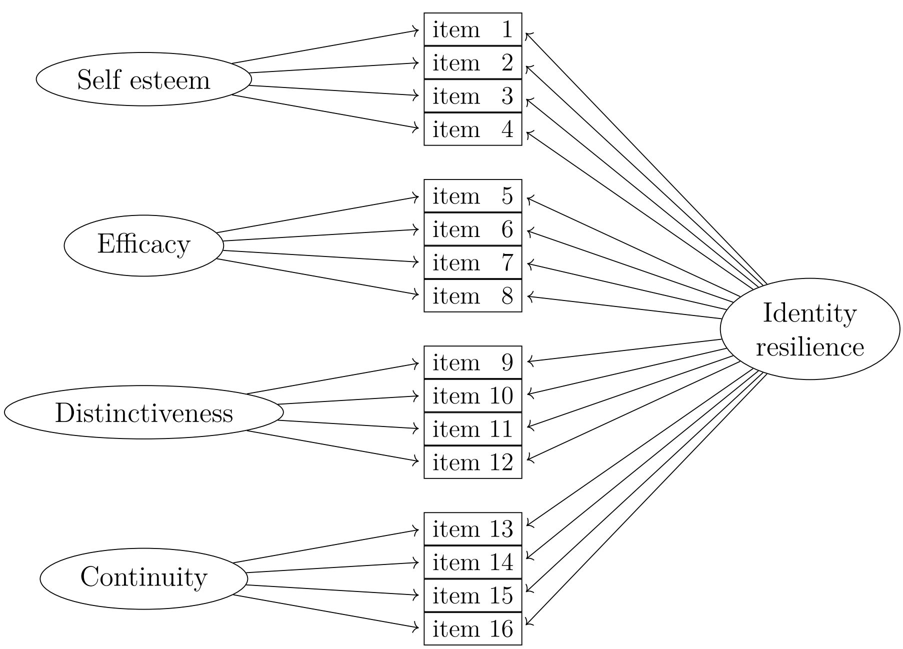

Figure 4. The bifactor plot for the Identity Resilience Scale.

research. Here a confirmatory approach is used, as both previous studies and theoretical analysis support this conceptualisation (Breakwell 2023).

Comparing scales by countries

To illustrate typical comparisons, here the values on these constructs (the single trust in science measure, the bifactor identify resilience and the four components) for the two countries are compared. Table 2 shows the means, their 95 per cent confidence intervals (CI), a t test comparing these and a common effect size for this comparison called Cohen's d (the difference in means divided by the SD). The p-values for the individual tests are printed as well as those after adjusting using Holm's method. Holm's method is used because having six t-tests means the probability of getting a significant result (i.e., $p < .05$) on at least one of these, even if there are no differences in the population, is much higher than 5 per cent (Bretz *et al.* 2010). An alternative way to consider these differences is to look at the effect sizes. Cohen (1992) describes $d = .20$ as a small effect ($d = .50$ as medium and $d = .80$ as large). From this, the effects are 'small' for trust in science and distinctiveness for UK respondents, with US respondents tending to score higher.

There are several assumptions of these t-tests, including that the within-group population distributions are normally distributed and with equal variances. These assumptions will never be correct: researchers should 'move from [the idea that] all assumptions are right towards all assumptions are wrong' (Tukey 1986: 72). This does not mean that they should be ignored, but that even with relatively small deviations the results will probably still enable wise decision-making. Another assumption is that the group variable is measured without error. This should not be an issue for these comparisons as people should accurately know which country they are in. It is also assumed that the data are independent of each other. This is another reason that only using a single respondent from each IP address is good practice. Two people from the same IP address are likely to be more similar to each other and while there are methods to take into account non-independence of data (e.g., Goldstein 2011), they would not be practical to apply for a small number of duplicate IP addresses.

The relationships among these constructs

In several of our articles in this special issue and elsewhere (e.g., Breakwell *et al.* 2023), and in papers by others using social surveys, researchers put forward a causal model for the relationships among variables, seeing how well the data fit the model and then focusing on the relationships between the pairs of constructs. This is called *path*

Table 2. Comparing the means for the United Kingdom and United States. Student *t*-tests with their associated *p*-values (without and with Holm's adjustment for the number of tests) are shown with Cohen's *d*. 95 per cent. CIs (confidence intervals) are shown below the means and *d*.

	$\bar{x}_{UK}$	$\bar{x}_{US}$	t	df	p	p_{adj}	Cohen's d
Trust in Science	−0.085	0.124	4.305	1,878	< .001	< .001	0.202
CI	(−0.145, −0.025)	(0.049, 0.199)					(0.110, 0.295)
Identity Resilience	0.034	−0.050	2.012	1,878	.044	.177	−0.095
CI	(−0.015, 0.083)	(−0.119, 0.019)					(−0.187, −0.002)
Self Esteem	0.018	−0.027	1.424	1,878	.155	.309	−0.067
CI	(−0.020, 0.056)	(−0.077, 0.023)					(−0.159, 0.025)
Efficacy	0.018	−0.027	1.216	1,878	.224	.309	−0.057
CI	(−0.025, 0.062)	(−0.087, 0.034)					(−0.149, 0.035)
Distinctiveness	−0.070	0.103	4.041	1,878	< .001	< .001	0.190
CI	(−0.122, −0.018)	(0.036, 0.170)					(0.098, 0.282)
Continuity	−0.030	0.044	1.856	1,878	.064	.191	0.087
CI	(−0.079, 0.020)	(−0.016, 0.103)					(−0.005, 0.179)

analysis. One popular approach is called *structural equation modelling* (SEM). There are a few approaches to this. Most frequently this involves simultaneously fitting a model that incorporates both the measurement of the latent constructs and the relations among them. This model could be run separately for the United Kingdom or United States, or with both to compare the effects for each country. This would allow the examination of differences by the constructs and the relationships among them, resulting in a complex model. Alternatively, a two-step approach (e.g., Anderson & Gerbing, 1988) where the measurement of the constructs occurs in the first step and the relationships among the constructs, or the structural part of the model, in the second is possible. This is also called the *structural after measurement* (SAM) approach. This means that constructs are constructed in the same way for the two countries. While there are other approaches that could be used, the SAM approach will be used for illustration. It is important to note that it is not always possible to conceptually separate the measurement and structural parts of a model.

One approach to this two-stage approach would be to estimate the constructs, as was done earlier in this article, and use these in a set of regression equations. As discussed with respect to *t*-tests, these would assume that the predictor variables are measured without error. Rosseel and Loh (in press) describes several problems with this approach but note that it is still popular. As with the correlations, it tends to underestimate the associations among the variables. The alternative is to include the uncertainty in these estimates in the model. Rosseel and Loh (in press) show the equations for doing this and provide a function *sam* in a new version of the package *lavaan* (Rosseel 2012). This is the approach used here. More variables are included here to show the approach and to match our research elsewhere in this issue.

SEM, and more generally using sets of regression models to create path diagrams, is sometimes referred to as causal modelling, as if causation can suddenly be determined from the correlations by using fancy statistics. Morgan and Winship (2015) note how some people blamed the over-reliance and over-optimism in these models for many negative consequences in the social sciences.

> Naive usage of regression modeling was blamed for nearly all the ills of sociology, everything from stripping temporality and context from the mainstream, … the suppression of attention to explanatory mechanisms, … the denial of causal complexity, … and the destruction of mathematical sociology. (Morgan & Winship 2015: 13)

Cartwright (2014: 308) describes the situation succinctly as 'no causes in, no causes out'. I assume identity resilience, social support and trust in science are related, and that trust in science influences COVID preventative behaviours, but our statistical procedures cannot show if the direction of causation is accurate. This is a framework in which to test our hypotheses about country differences. Country is treated as exogenous, and none of these other variables will influence it appreciably (there may be some influence, for example, someone who really trusts science might feel compelled to move to the United Kingdom, but this influence will be small enough for us to ignore). Our interest is in whether country influences trust in science and whether, after accounting for the influence of trust in science on COVID behaviours, country further influences COVID behaviours. It does. This can be shown by comparing the model in Figure 5 without the dashed line with the model with the dashed line. The difference in fit is: $\chi^2(1) = 39.72$, $p < .001$. This suggests this effect (the dashed arrow of Figure 5) should be included in the model.

Like the procedure itself, the numeric results from the SAM model are separated into measurement and structural parts. The measurement part estimates the reliability for each construct. These are (for the model, including the dashed line in Figure 5):

Identity Resilience	.901
Trust in Science	.927
Social Support	.915
COVID Prevention	.853

Statistics related to the structural aspects of the model are shown in Table 3. The largest effects are for the identity resilience to social support edge and the trust in science to COVID prevention edge. Identifying these is important for understanding the

Table 3. Path coefficients and related statistics for the model shown in Figure 5.

Path	Coef.	*se*	*z*	*p*
Identity Resilience → Social Support	−0.551	0.026	−21.546	< .001
Identity Resilience → Trust in Science	−0.084	0.023	−3.726	< .001
Country → Trust in Science	−0.154	0.035	−4.437	< .001
Trust in Science → COVID Prevention	−0.484	0.038	−12.772	< .001
Country → COVID Prevention	0.275	0.048	5.704	< .001

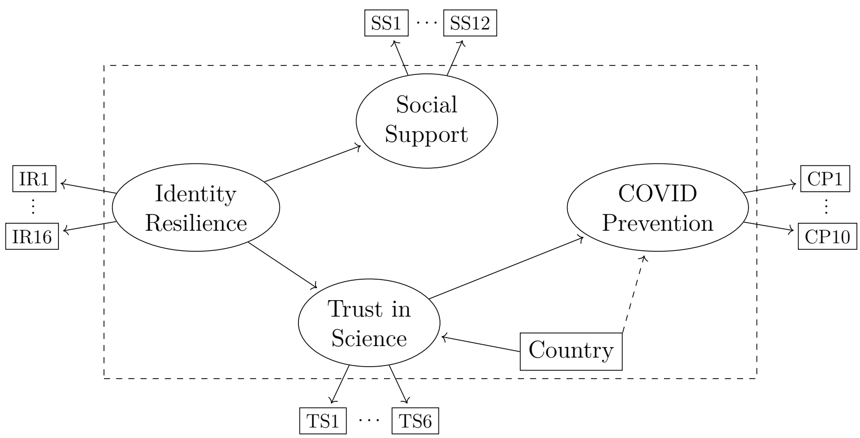

Figure 5. A SAM model (Rosseel & Loh, in press) of the relationships among COVID-19-related constructs.

Note: The observed variables that make up the constructs and their error terms are not shown. The dashed rectangle encloses the structural part of the model. The measurement is shown by the paths between the constructs inside the rectangle and the items outside it. Each of the observed items, other than country, also has an implied error term associated with it, but these are not shown as the figure already is fairly complex. The coefficients for this model are shown in Table 3.

relationships among these constructs and related behaviours like having the vaccine. In particular, the path between trust in science and preventative behaviours suggests increasing trust in science may improve behaviours that in turn may lessen the impact of pandemics.

Summary

Considerations and assumptions are part of all research, even when not evident in the descriptions of the research in scientific journals. The aim of this paper, like Toto, is to pull back the curtain that obscures how the procedures work, but unlike the film *Wizard of Oz* where the 'wizard' did not want people to see what occurred behind the curtain, I want you to pay attention to those considerations and assumptions behind the curtain of scientific protocols. Word limits, blind reviews and years of training have led social scientists to report the main findings from research in a succinct and formal manner that obfuscates decisions made when conducting the research. The format of journals and fifteen-minute conference slots make this inevitable and we do not believe the approach used here is appropriate for all dissemination. For this special issue, focused on several of our research ideas and written for a broad audience, it is worth describing our considerations and assumptions in more detail. Further, reflecting on these helps to focus on these decisions and forced me to make the reasons why certain choices were made explicit.

We created a research team and developed research questions in response to a call from the British Academy. Our aim was to explore different aspects of IPT in the United Kingdom and United States with focus on different ethnicities within the context of COVID-19. Our main method, for much of our research, has been to present a set of scales to survey respondents and draw conclusions about how people think and behave based on the associations among their responses. This is a tall order and requires both assumptions and some empirical checks of *some of* these assumptions. We assume that the sample achieved online through Prolific will be similar enough to others to provide useful and meaningful results. We assume that participants' responses inform us about their beliefs and behaviours consistent with our intent. We assume these can be aggregated and represent the intended psychological constructs. The choice of statistical methods for this aggregation and for looking at the associations among the constructs also require decisions. In the typical article, the authors focus more on *what* they did rather than *why* they did what they did, and *why* they didn't do the alternatives.

Social science theory and methods can help inform policy and other applications related to societally important issues. The COVID-19 pandemic is an example. While medical and economic research are vital for pandemics, so is understanding how people will psychologically react to health guidance and restrictions. Social scientists have many tools at their disposal. When faced with a global crisis the research tools from many disciplines can be useful. Each discipline has tacit protocols. Being explicit about the protocols helps readers to better understand the approaches and the research implications.

Acknowledgements

This research was funded by the British Academy (CRUSA210025 to JR, GB, JB and DW); thanks to my colleagues for discussion on these topics and on this paper.
Correspondence should be sent to daniel.wright@unlv.edu or dbrookswr@gmail.com,

or Educational Psychology, University of Nevada, Las Vegas, Box 453001, 4505 S. Maryland Pkwy., Las Vegas, NV 89154-3001, USA.

References

Anderson, J. C. & Gerbing, D. W. (1988), 'Structural equation modeling in practice: a review and recommended two-step approach', *Psychological Bulletin*, 103: 411–423. https://doi.org/10.1037/0033-2909.103.3.411

Auerswald, M. & Moshagen, M. (2019), 'How to determine the number of factors to retain in exploratory factor analysis: a comparison of extraction methods under realistic conditions', *Psychological Methods*, 24: 468–491. https://doi.org/10.1037/ met0000200

Bacon, F. (2019 [1620]), *Novum organum* (Anodos Books).

Bacon, F. (2020 [1626]), *New Atlantis* (Dallas, TX, CTMS Publishers).

Bartholomew, D. J. (1993), 'Estimating relationships between latent variables', *Sankhyā: The Indian Journal of Statistics*, Series A, 55: 409–419. https://doi.org/10.2307/25050951

Bartholomew, D. J., Knott, M. & Moustaki, I. (2011), *Latent Variable Models and Factor Analysis: A Unified Approach* (3rd edn; Chichester, Wiley). https://doi.org/10.1002/9781119970583

Biffignandi, S. & Bethlehem, J. (2021), *Handbook of Web Surveys* (2nd edn; Hoboken, NJ, Wiley). https://doi.org/10.1002/9781119371717

Breakwell, G. M. (2023), *Identity: Unique and Shared* (London, SAGE).

Breakwell, G. M., Fino, E. & Jaspal, R. (2022), 'The Identity Resilience Index: development and validation in two UK samples', *Identity*, 22: 166–182. https://doi.org/10.1080/15283488.2021.1957895

Breakwell, G. M., Jaspal, R. & Wright, D. B. (2023), 'Identity resilience, science mistrust, COVID-19 risk and fear predictors of vaccine positivity and vaccination likelihood: a survey of UK and Portuguese samples', *Journal of Health Psychology*, 8: 747–759. https://doi.org/10.1177/13591053231161891

Bretz, F., Hothorn, T. & Westfall, P. (2010), *Multiple Comparisons Using R* (New York, CRC Press).

Campbell, N. R. (1928), 'Personal and impersonal styles in scientific communication', *Nature*, 121: 1021. https://doi.org/10.1038/1211021a0

Cartwright, N. (2014), 'Causal inference', in Cartwright, N. & Montuschi, E. (eds), *Philosophy of Social Science* (Oxford, Oxford University Press), 308–326.

Cattell, R. B. (1966), 'The scree test for the number of factors', *Multivariate Behavioral Research*, 1: 245–276. https://doi.org/10.1207/s15327906mbr0102\10

Chambers, J. M. (2008), *Software for Data Analysis: Analysis with R* (New York, Springer). https://doi.org/10.1007/978-0-387-75936-4

Chambers, J. M. (2009), 'Facets of R', *R Journal*, 1(1): 5–8. https://journal.r-project.org/archive/2009-1/RJournal_2009-1_Chambers.pdf

Cohen, J. (1992), 'A power primer', *Psychological Bulletin*, 112: 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cronbach, L. J. (1951), 'Coefficient alpha and the internal structure of tests', *Psychometrika*, 16: 297–334. https://doi.org/10.1007/BF02310555

Feynman, R. P. (1974), 'Cargo cult science: some remarks on science, pseudoscience, and learning how to not fool yourself. Caltech's 1974 commencement address', *Engineering and Science*, 37(7): 10–13.

Gigerenzer, G. (1991), 'From tools to theories: a heuristic of discovery in cognitive psychology', *Psychological Review*, 98: 254–267. https://doi.org/10.1037/0033-295X.98.2.254
Goldstein, H. (2011). *Multilevel Statistical Models* (4th edn.; Chichester, Wiley). https://doi.org/10.1002/9780470973394
Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998), 'Measuring individual differences in implicit cognition: the implicit association test', *Journal of Personality and Social Psychology*, 74: 1464–1480. https://doi.org/10.1037/ 0022-3514.74.6.1464
Gummer, T., Roßmann, J. & Silber, H. (2021), 'Using instructed response items as attention checks in web surveys: properties and implementation', *Sociological Methods & Research*, 50: 238–264. https://doi.org/10.1177/0049124118769083
Guttman, L. (1945), 'A basis for analyzing test-retest reliability', *Psychometrika*, 10: 255–282. https://doi.org/10.1007/BF02288892
Kalton, G. (2021), *Introduction to Survey Sampling* (2nd edn.; London, SAGE). https://doi.org/10.4135/9781071909812
Loehlin, J. C. & Beaujean, A. A. (2017), *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis* (5th edn.; New York, Routledge). https://doi.org/10.4324/9781315643199
Luce, R. D. (1986), *Response Times: Their Role in Inferring Elementary Mental Organization* (Oxford, Oxford University Press). https://doi.org/10.1093/acprof:oso/9780195070019.001.0001
McNeish, D. (2018), 'Thanks coefficient alpha, we'll take it from here', *Psychological Methods*, 23: 412–433. https://doi.org/10.1037/met0000144
McNeish, D. & Wolf, M. G. (2020), 'Thinking twice about sum scores', *Behavior Research Methods*, 52: 2287–2305. https://doi.org/10.3758/s13428-020-01398-0
Mair, P. (2016), 'Thou shalt be reproducible! A technology perspective', *Frontiers in Psychology*, 7. https://doi.org/10.3389/fpsyg.2016.01079
Mizumoto, A. & Plonsky, L. (2016), 'R as a *lingua franca*: advantages of using R for quantitative research in applied linguistics', *Applied Linguistics*, 37: 284–291. https://doi.org/10.1093/applin/amv025
Morgan, S. L. & Winship, C. (2015), *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (2nd edn.; Cambridge, Cambridge University Press). https://doi.org/10.1017/CBO9781107587991
Muliak, S. A. (2010), *Foundations of Factor Analysis* (2nd edn., New York, CRC Press).
Nadelson, L., Jorcyk, C., Yang, D., Jarratt Smith, M., Matson, S., Cornell, K. & Husting, V. (2014), 'I just don't trust them: the development and validation of an assessment instrument to measure trust in science and scientists', *School Science and Mathematics*, 114: 76–86. https://doi.org/10.1111/ssm.12051
National Research Council. (2013), *Nonresponse in Social Science Surveys: A Research Agenda* (Washington, DC, National Academies Press). https://doi.org/10.17226/18293
Nisbett, R. E. & Wilson, T. D. (1977), 'Telling more than we can know: verbal reports on mental processes', *Psychological Review*, 84: 231–259. https://doi.org/10.1037/ 0033-295X.84.3.231
Peer, E., Rothschild, D., Gordon, A., Evernden, Z. & Damer, E. (2022), 'Data quality of platforms and panels for online behavioral research', *Behavior Research Methods*, 54: 1643–1662. https://doi.org/10.3758/s13428-021-01694-3
Popper, K. (1994), *The Myth of the Framework: In Defense of Science and Rationality* (London, Routledge).
R Core Team. (2022), *R: A Language and Environment for Statistical Computing* [computer software manual]. https://cran.r-project.org/doc/manuals/R-intro.pdf
Rosseel, Y. (2012), 'lavaan: an R package for structural equation modeling', *Journal of Statistical Software*, 48(2): 1–36. http://www.jstatsoft.org/v48/i02/

Rosseel, Y. & Loh, W. W. (in press), 'A structural after measurement approach to structural equation modeling', *Psychological Methods*. https://doi.org/10.1037/met0000503
Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys* (New York, John Wiley & Sons). https://doi.org/10.1002/9780470316696
Sheehan, K. B. & Pittman, M. (2016), *Amazon's Mechanical Turk for Academics: The HIT Handbook for Social Science Research* (Irvine, CA, Melvin & Leigh).
Spearman, C. (1904), '"General intelligence", objectively determined and measured', *American Journal of Psychology*, 15, 201–293. https://doi.org/10.2307/1412107
Thompson, B. (2003), 'Understanding reliability and coefficient alpha, really', in Thompson, B. (ed.), *Score Reliability* (Los Angeles, SAGE), 3–30. https://doi.org/10.4135/9781412985789.n1
Tukey, J. W. (1986), 'Sunset salvo', *The American Statistician*, 40: 72–76. https://doi.org/10.1080/00031305.1986.10475361
Van Buuren, S. (2018), *Flexible Imputation of Missing Data* (2nd edn.; New York, Chapman & Hall/CRC Press). https://doi.org/10.1201/9780429492259
Velicer, W. F., Eaton, C. A. & Fava, J. L. (2000), 'Construct explication through factor or component analysis: a review and evaluation of alternative procedures for determining the number of factors or components', in Goffin, R. D. & Helmes, E. (eds), *Problems and Solutions in Human Assessment* (Norwell, MA, Kluwer), 41–71. https://doi.org/10.1007/978-1-4615-4397-8_3
Waller, N. G. & Meehl, P. E. (1998), *Multivariate Taxometric Procedures: Distinguishing Types from Continua* (Thousand Oaks, CA, SAGE).
Widaman, K. F. & Revelle, W. (2022), 'Thinking thrice about sum scores, and then some more about measurement and analysis', *Behavior Research Methods*, 55: 788–806. https://doi.org/10.3758/s13428-022-01849-w
Wise, S. L. (2017), 'Rapid-guessing behavior: its identification, interpretation, and implications', *Educational Measurement: Issues and Practice*, 36: 52–61. https://doi.org/10.1111/emip.12165
Wise, S. L. & Kong, X. (2005), 'Response time effort: a new measure of examinee motivation in computer-based tests', *Applied Measurement in Education*, 18, 163–183. https://doi.org/10.1207/s15324818ame1802_2
Wolff, S. M., Breakwell, G. M. & Wright, D. B. (forthcoming), *Psychometric Evaluation of the Trust in Science and Scientists Scale*.
Wright, D. B. & Wells, S. M. (2020), 'Creating latent variables', in Breakwell. G. M., Wright, D. B. & Barnett, J. (eds), *Research Methods in Psychology* (5th edn.; London, SAGE Publications).
Wright, D. B., Wolff, S. M., Jaspal, R., Barnett, J. & Breakwell, G. M. (2022), 'The choice of response alternatives in COVID-19 social science surveys', *PLoS One*, 17(11): e0263552. https://doi.org/10.1371/journal.pone.0263552
Xie, Y. (2015), *Dynamic Documents with R and knitr* (2nd edn.; New York, Chapman & Hall/CRC). https://doi.org/10.1201/b15166

To cite the article: Wright, D. B. (2023), 'Methodological considerations and assumptions in social science survey research', *Journal of the British Academy*, 11(s5): 35–58.
https://doi.org/10.5871/jba/011s5.035

Journal of the British Academy (ISSN 2052–7217) is published by
The British Academy, 10–11 Carlton House Terrace, London, SW1Y 5AH
www.thebritishacademy.ac.uk